

GROUPS RELATED TO GAUSSIAN GRAPHICAL MODELS

JAN DRAISMA, SONJA KUHNT, AND PIOTR ZWIERNIK

ABSTRACT. Gaussian graphical models have become a well-recognized tool for the analysis of conditional independencies within a set of continuous random variables. We describe the maximal group acting on the space of covariance matrices, which leaves this model invariant for any fixed graph. This group furnishes a representation of Gaussian graphical models as composite transformation families and enables to analyze properties of parameter estimators. We use these results in the robustness analysis to compute upper bounds on finite sample breakdown points of equivariant estimators of the covariance matrix. In addition we provide conditions on the sample size so that an equivariant estimator exists with probability 1.

1. INTRODUCTION

Using Gaussian graphical models is a popular way of modelling complex associations in the multivariate continuous case. Statistical inference for this model class often depends on the maximum likelihood estimator of the covariance matrix. As pointed out for example by Maronna, Martin, Yohai [16] the maximum likelihood estimator is typically the least robust with respect to potential outliers in the sample space. For this reason it may be interesting to understand properties of more robust estimators for this model class.

Robustness issues with respect to graphical models have been rarely looked at so far, although it has been known for some time that the classical estimators and model selection procedures are vulnerable to contaminated data [11][14]. First approaches towards robust covariance estimators for a given Gaussian model with undirected graph can be found in Becker [5] and Gottard, Pacillo [12]. The authors suggested there to replace the sample covariance matrix by the reweighted MCD estimator. Miyamura and Kano [18] propose to use an M-type estimator instead. Finegold and Drton [9] as well as Vogel and Fried [23] discard the assumption of normality and consider the t-distribution and general elliptical distribution, respectively, to model heavy tails. Ravikumar et al. [19] as well as Meinshausen and Bühlmann [17] deal with the question of a robust selection of the graph structure by L_1 -regularization or penalization.

We use symmetries of the Gaussian graphical model in order to understand the behavior of an arbitrary equivariant estimator of the covariance matrix. Having an explicit group acting on a statistical model has numerous advantages. This was first pointed out by Fisher in the context of the location and scale models [10], which then led to the definition of

The first and third author are supported by Draisma's Vidi grant from the Netherlands Organisation for Scientific Research (NWO).

transformation families [4], [20, Section 6.2]. A basic idea behind the transformation family in statistics is that of a statistical model which is invariant under the action of a group and this action is assumed transitive. This representation can be then used for example in the study of model invariants and distributional aspects of the maximum likelihood estimator and other equivariant estimators.

In the present paper we use the symmetries of Gaussian graphical models in the robustness analysis for equivariant estimators of the covariance matrix, based on a more general theory developed by Davies and Gather [8]. This concerns an important measure of robustness for parameter estimators, which is the high breakdown point (see for example [16]). Intuitively, the high breakdown point tells us what proportion of the data can be changed arbitrarily without arbitrarily large changes in the value of a given estimator. In this paper we work out explicit upper bounds on breakdown points of equivariant estimators. These bounds depend only on the graph \mathcal{G} . We introduce these results in Section 2. This also includes one of the three main results of this paper, which we state in Theorem 2.6.

The group of interest can be defined as follows. Let $\mathbf{X} = (X_i)_{i \in [m]}$ be a random vector with multivariate normal distribution $\mathcal{N}(0, \Sigma)$. We assume the zero mean for simplicity. Denote by \mathcal{S}_m the set of symmetric matrices in $\mathbb{R}^{m \times m}$ and by \mathcal{S}_m^+ its subset of positive definite matrices. Let $\mathcal{G} = ([m], E)$ be an undirected graph with set of vertices $[m] := \{1, \dots, m\}$ and set of edges $E \subseteq [m] \times [m]$. By $\mathcal{S}_{\mathcal{G}} \subseteq \mathcal{S}_m$ denote the linear space of symmetric matrices whose (i, j) off-diagonal entry is zero if $(i, j) \notin E$, and by $\mathcal{S}_{\mathcal{G}}^+$ the subset of $\mathcal{S}_{\mathcal{G}}$ given by all positive definite matrices. The *Gaussian graphical model* is the statistical model for \mathbf{X} given by

$$M(\mathcal{G}) = \{\mathcal{N}(0, \Sigma) : \Sigma^{-1} \in \mathcal{S}_{\mathcal{G}}^+\}.$$

Hence the expression for the normal log-density is a quadratic polynomial in x_1, \dots, x_m given by

$$\log f(\mathbf{x}; K) = -\frac{m}{2} \log 2\pi + \frac{1}{2} \det K - \frac{1}{2} \sum_{i \in [m]} K_{ii} x_i^2 - \sum_{(i,j) \in E} K_{ij} x_i x_j,$$

where $K := \Sigma^{-1}$ is the concentration matrix.

The general linear group $\mathrm{GL}_m(\mathbb{R})$ acts on the sample space \mathbb{R}^m by the matrix multiplication

$$g \cdot \mathbf{x} := g\mathbf{x}$$

for $g \in \mathrm{GL}_m(\mathbb{R})$. Since $f(\mathbf{x}; K) = f(g\mathbf{x}; g^{-T}Kg^{-1})$, the action of $\mathrm{GL}(\mathbb{R}^m)$ on \mathbb{R}^m induces the action on \mathcal{S}_m^+ given by

$$g \cdot K := g^{-T}Kg^{-1}.$$

In this paper we are interested in the following problem.

- (1) Find all $g \in \mathrm{GL}_m(\mathbb{R})$ such that $g \cdot \mathcal{S}_{\mathcal{G}} = \mathcal{S}_{\mathcal{G}}$.

Since in this case also $g \cdot \mathcal{S}_{\mathcal{G}}^+ = \mathcal{S}_{\mathcal{G}}^+$, we say that g leaves $M(\mathcal{G})$ invariant (see [20, Definition 6.27]); the set of all such g is a subgroup of $\mathrm{GL}_m(\mathbb{R})$ denoted by G .

The main concept of the paper is that of a G -equivariant estimator. Let $\mathbf{x}_{(n)} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ be a random sample and $(\mathbb{R}^m)^n$ the space of all random samples of size n . An estimator is by definition a map $T_n : (\mathbb{R}^m)^n \rightarrow \mathcal{S}_{\mathcal{G}}^+$, which is defined almost everywhere. Let

$$g \cdot \mathbf{x}_{(n)} := (g\mathbf{x}_1, \dots, g\mathbf{x}_n).$$

We say that an estimator T_n is G -equivariant if $T_n(g \cdot \mathbf{x}_{(n)}) = g \cdot T_n(\mathbf{x}_{(n)})$ for every $\mathbf{x}_{(n)} \in (\mathbb{R}^m)^n$ and $g \in G$ whenever $T_n(\mathbf{x}_{(n)})$ exists. The equivariance is an important property for estimators as it assures that the estimator is stable under simultaneous compatible transformations of the parameter and the sample space (see for example [20, Chapter 6]).

Our main results can be summarized as follows. For a fixed undirected graph $\mathcal{G} = ([m], E)$ define a preorder on $[m]$ by

$$(2) \quad i \preceq j \quad \text{if} \quad N(j) \cup \{j\} \subseteq N(i) \cup \{i\},$$

where $N(i)$ denotes the set of neighbors of i in \mathcal{G} . Consider the set G^0 of all matrices in $\text{GL}_m(\mathbb{R})$ such that $g_{ij} \neq 0$ only if $j \preceq i$. We show later in Section 3.2 that this set forms a subgroup of $\text{GL}_m(\mathbb{R})$. The preorder defined in (2) gives an equivalence relation on $[m]$ by $i \sim j$ if and only if $i \preceq j$ and $j \preceq i$. Let now $\tilde{\mathcal{G}}$ be the graph on the set of equivalence classes $[m]/\sim$ induced by \mathcal{G} by identifying all equivalent nodes. Hence, two equivalence classes \bar{i} and \bar{j} are connected by an edge in $\tilde{\mathcal{G}}$ if i and j are neighbors in \mathcal{G} . We color the vertices of $\tilde{\mathcal{G}}$ by positive numbers representing cardinality of the corresponding equivalence class. The automorphisms of $\tilde{\mathcal{G}}$, denoted by $\text{Aut}(\tilde{\mathcal{G}}, c)$, are the automorphisms of the underlying graph preserving the coloring of the vertices.

Theorem 1.1. *For every undirected graph $\mathcal{G} = ([m], E)$ the group G of elements satisfying (1) is generated by its subgroups G^0 and $\text{Aut}(\mathcal{G})$. More precisely $G = G^0 \rtimes \text{Aut}(\tilde{\mathcal{G}}, c)$, that is G is a semidirect product of G^0 and $\text{Aut}(\tilde{\mathcal{G}}, c)$.*

A natural question is what are the conditions on n and \mathcal{G} so that the maximum likelihood estimator (MLE) exists with probability one. For a discussion see the recent paper [22] and references therein. Another related question is, what are the conditions for n and \mathcal{G} so that there exists any G -equivariant map $T_n : (\mathbb{R}^m)^n \rightarrow \mathcal{S}_{\mathcal{G}}^+$, $\mathbf{x}_{(n)} \mapsto T_n(\mathbf{x}_{(n)})$. Since the MLE is G -equivariant, these conditions are weaker than the conditions for existence of the MLE.

Theorem 1.2. *Let $\mathcal{G} = ([m], E)$ be an undirected graph. There exists a G -equivariant estimator $T_n : (\mathbb{R}^m)^n \rightarrow \mathcal{S}_{\mathcal{G}}^+$ if and only $n \geq \max_{i \in [m]} |\{j \in [m] : j \preceq i\}|$.*

The paper is organized as follows. In Section 2 we formally introduce the concept of a breakdown point. We also link the computation of breakdown points to the symmetry of a statistical model. We also state Theorem 2.6 which gives the formula for the finite sample breakdown point of any G -equivariant estimator of the covariance matrix. In Section 3 we introduce various combinatorial structures associated to Gaussian graphical models. These will be then used in the analysis of the connected component of the identity of the group G defined in (1). In Section 4 we complete our description of the group G , which is given

in Theorem 1.1. Examples of various graphs and corresponding groups G are provided in Section 5. In Section 6 we analyze G -equivariant estimators more closely, which leads to the proof of Theorem 1.2 and Theorem 2.6. In Section 7 we complete our geometric description of the group G by a study of G -orbits of generic concentration matrices in \mathcal{S}_G^+ . In particular Theorem 7.2 gives us as a result the dimension of a generic G -orbit. We conclude the paper with a short discussion.

2. BREAKDOWN POINTS AND GROUPS

An important concept to measure the robustness of parameter estimators are breakdown points. In a simple univariate situation, if the estimator is given by the sample mean, then an arbitrarily large change made to one of the observations leads to an arbitrarily large change in the value of the estimator. If we consider the sample median, then changing one observation in a sample of size larger than two cannot lead to arbitrarily large changes in the estimator. This feature makes the median more robust to outliers in the sample. Informally speaking a finite sample breakdown point is the minimal number of sample points such that altering them effects in arbitrarily large changes in the value of an estimator. This quantity is usually normalized by the sample size n (see also Definitions 2.2 and 2.3). So for example in the case of the sample mean above, the breakdown point is $1/n$. In terms of robustness the estimator with the highest breakdown point is preferred [13].

We now briefly discuss the main results of [8], which link the concept of a breakdown point and that of symmetry groups of statistical models. This is the main motivation for the study of the group structure of Gaussian graphical models. To formally define the breakdown point in the general case we need to introduce more concepts.

A *pseudometric* on the parameter space \mathcal{S}_G^+ is a function $D : \mathcal{S}_G^+ \times \mathcal{S}_G^+ \rightarrow \mathbb{R}_{\geq 0}$ such that for every $K, L, M \in \mathcal{S}_G^+$:

1. $D(K, K) = 0$,
2. $D(K, L) = D(L, K)$,
3. $D(K, M) \leq D(K, L) + D(L, M)$.

The only difference between a metric and a pseudometric is that we may have $D(K, L) = 0$ even if $K \neq L$. In this paper we work with a special type of a pseudometric on \mathcal{S}_G^+ given by

$$(3) \quad D(K_1, K_2) = |\log \det(K_1 K_2^{-1})|.$$

The following example illustrates the main idea of the link between the breakdown of an equivariant estimator to symmetry groups.

Example 2.1. Let G be a subgroup of $\text{GL}_m(\mathbb{R})$ and let G_1 be the set of $g \in G$ such that $\det(g^T g) \neq 1$. For a random sample $\mathbf{x}_{(n)}$ assume that $g \in G_1$ stabilizes the first $n - k$ sample points but alters the remaining k points, that is $g\mathbf{x}_i = \mathbf{x}_i$ for $i = 1, \dots, n - k$ and $g\mathbf{x}_i \neq \mathbf{x}_i$ for $i = k + 1, \dots, n$. Then for any $l \geq 1$ we have $g^l \mathbf{x}_i = \mathbf{x}_i$ if $i = 1, \dots, n - k$. Hence $(g^l \cdot \mathbf{x}_{(n)})_{l \geq 1}$ is a sequence of data samples which all differ from $\mathbf{x}_{(n)}$ in k data points.

We note that if T_n is a G -equivariant estimator then for D defined in (3)

$$D(T_n(\mathbf{x}_{(n)}), T_n(g^l \cdot \mathbf{x}_{(n)})) = D(T_n(\mathbf{x}_{(n)}), g^l \cdot T_n(\mathbf{x}_{(n)})) = l |\log \det(g^T g)| \xrightarrow{l \rightarrow \infty} \infty.$$

In this way we have constructed a sequence of data samples with $n - k$ points fixed such that the resulting estimator can be arbitrarily far from the original estimator $T_n(\mathbf{x}_{(n)})$ with respect to the pseudometric D .

Consider now the measurable sample space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, where $\mathcal{B}(\mathbb{R}^m)$ is the Borel σ -algebra, and the family \mathfrak{P} of all probability measures nondegenerate with respect to the Lebesgue measure on this space. For any $g \in G$ and $P \in \mathfrak{P}$ we define $P^g := g \cdot P$ by $P^g(B) = P(g^{-1}(B))$. This is consistent with the action of G on \mathbb{R}^m .

A map $T : \mathfrak{P} \rightarrow \mathcal{S}_G^+$ is called *equivariant* if $T(P^g) = g \cdot T(P)$ for every $g \in G$. We assume that a pseudometric d is defined on \mathfrak{P} which satisfies

$$\sup_{P, Q \in \mathfrak{P}} d(P, Q) = 1$$

and for all $P, Q_1, Q_2 \in \mathfrak{P}$ and $\alpha, 0 < \alpha < 1$,

$$d(\alpha P + (1 - \alpha)Q_1, \alpha P + (1 - \alpha)Q_2) \leq 1 - \alpha.$$

We also consider a pseudometric $D : \mathcal{S}_G^+ \times \mathcal{S}_G^+ \rightarrow \mathbb{R}$ that satisfies $\sup_{K_1, K_2} D(K_1, K_2) = \infty$. In this paper we constrain ourselves solely to the pseudometric in (3), but other distances can be used as well.

Definition 2.2. The breakdown point $\epsilon^*(T, P, d, D)$ of T at the distribution P with respect to pseudometrics d and D is defined by

$$(4) \quad \epsilon^*(T, P, d, D) = \inf\{\epsilon > 0 : \sup_{d(P, Q) < \epsilon} D(T(P), T(Q)) = \infty\}.$$

We can also define a breakdown point for an estimator $T_n : (\mathbb{R}^m)^n \rightarrow \mathcal{S}_G^+$. If $\mathbf{x}_{(n)} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a sample of size n , then by $\mathbf{y}_{n,k}$ denote a sample obtained from $\mathbf{x}_{(n)}$ by altering at most k of the \mathbf{x}_i .

Definition 2.3. The *finite sample breakdown point* of a G -equivariant estimator T_n at the sample $\mathbf{x}_{(n)}$ is defined by

$$(5) \quad \text{fsbp}(T_n, \mathbf{x}_{(n)}, D) = \frac{1}{n} \min\{k \in [n] : \sup_{\mathbf{y}_{n,k}} D(T_n(\mathbf{x}_{(n)}), T_n(\mathbf{y}_{n,k})) = \infty\}.$$

In particular $\text{fsbp}(T_n, \mathbf{x}_{(n)}, D) \geq 1/n$.

Intuitively Example 2.1 shows that symmetries of a statistical model can be exploited in order to construct a sequence of data points such that the supremum above is infinite. Now we explain formally how the group G is used in (7) to find upper bounds for $\epsilon^*(T, P, d, D)$ and $\text{fsbp}(T_n, \mathbf{x}_{(n)}, D)$ for every equivariant functional T . We need

$$G_1 = \{g \in G : \liminf_{l \rightarrow \infty} D(K, g^l \cdot K) = \infty\}.$$

Let I denote the identity matrix in G and $g|_B$ the restriction of $g \in G$ to a set $B \in \mathcal{B}(\mathbb{R}^m)$. Of importance for the upper breakdown bound is

$$(6) \quad \Delta(P) = \sup\{P(B) : B \in \mathcal{B}(\mathbb{R}^m), g|_B = I|_B \text{ for some } g \in G_1\},$$

where in the finite sample case P will be the sample distribution $P_{\mathbf{x}_{(n)}} := \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{x}_i}$ and otherwise $P \in \mathfrak{P}$.

Theorem 2.4 (Davies, Gather[8]). *With the above notation and under the assumption that $G_1 \neq \emptyset$ we have*

$$(7) \quad \epsilon^*(T, P, d, D) \leq \frac{1 - \Delta(P)}{2}.$$

For the finite-sample breakdown point we have

$$(8) \quad \text{fsbp}(T_n, \mathbf{x}_{(n)}, D) \leq \frac{1}{n} \left\lfloor \frac{n - n\Delta(P_{\mathbf{x}_{(n)}}) + 1}{2} \right\rfloor.$$

If D is defined by (3) then G_1 is the set of $g \in G$ such that $\det gg^T \neq 1$. Since g in Example 2.1 fixes the first $n - k$ data points in $\mathbf{x}_{(n)}$ and alters the remaining k points, we can construct a sequence $g^l \mathbf{x}_{(n)}$ for $l \geq 1$. This shows that by changing k data points arbitrarily we can make the new estimated value arbitrarily far from $T_n(\mathbf{x}_{(n)})$ and hence $\text{fsbp}(T_n, \mathbf{x}_{(n)}, D) < k/n$. Note that this construction leads to a worse bound than in the above theorem. However, a more efficient construction giving the bound in (8) has a similar flavor (see [7, Proof of Theorem 2.1]).

In [8, Section 4.2] Davies and Gather discuss the following example.

Example 2.5. (Scatter functionals and the affine group) With the sample space \mathbb{R}^m let the parameter space be given by \mathcal{S}_m^+ with the natural action of the affine group $g(x) = Ax + b$ inducing the action on Θ . With the pseudometric in (3) we have

$$\Delta(P) = \sup\{P(B) : B \text{ is a linear subspace of dimension } \leq m - 1\}.$$

If $P \in \mathfrak{P}$ then $\Delta(P) = 0$, but plugging in sample distributions $P_{\mathbf{x}_{(n)}}$ leads to positive values of $\Delta(P_{\mathbf{x}_{(n)}})$.

Note that this example includes the Gaussian graphical model for the full graph K_m . In this paper we generalize this for any graph \mathcal{G} . We are looking for elements in G_1 stabilizing some $B \in \mathcal{B}(\mathbb{R}^n)$. The set $B = \{\mathbf{x} : g\mathbf{x} = \mathbf{x} \text{ for some } g \in G_1\}$ forms a linear subspace of \mathbb{R}^m of positive codimension. Hence, if $P \in \mathfrak{P}$ then for any such $B \in \mathcal{B}(\mathbb{R}^m)$ the measure $P(B)$ of B is zero. Hence $\Delta(P) = 0$ and Theorem 2.4 now states that $\epsilon^*(T, P, d, D) \leq 1/2$ for any G -equivariant map T on \mathfrak{P} .

In the remaining of this paper we are primarily interested in the finite sample case, where $\mathbf{x}_{(n)} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a random sample from the model. The following result is the third main theorem of the paper.

Theorem 2.6. *Let $\mathcal{G} = ([m], E)$ be an undirected graph. Then*

$$\Delta(P) = \sup\{P(B) : B \text{ is a linear subspace of dimension } \leq Q - 1\},$$

where $Q = \max_{i \in [m]} |\{j \in [m] : j \preccurlyeq i\}|$. Moreover, if $\mathbf{x}_{(n)}$ is a generic sample (see Definition 6.5) then

$$(9) \quad \text{fsbp}(T_n, \mathbf{x}_{(n)}, D) \leq \frac{1}{n} \lfloor \frac{n - Q + 2}{2} \rfloor$$

for any G -equivariant estimator T_n .

Note that if a G -equivariant map exists then by Theorem 1.2 $n \geq Q$ and hence the right-hand side of (9) is always at least $1/n$. If $n - Q \in \{0, 1\}$ then the lower and the upper bound on the finite sample breakdown point of T_n coincide. The upper bound on Q is m and it is attained only for the complete graph K_m . The lower bound $Q = 1$ is attained for example for m -cycles where $m \geq 4$ and for graphs with no edges.

Remark 2.7. If \mathcal{G} is an undirected tree then $Q = 2$. From this it follows that for generic data sets $\text{fsbp}(T_n, \mathbf{x}_{(n)}, D) \leq \frac{1}{n} \lfloor \frac{n}{2} \rfloor$. More generally whenever Q is small, like in the case of trees, the upper bound for the finite sample breakdown point is close to $\frac{1}{n} \lfloor n/2 \rfloor$. On the other hand if T_n is the MLE then the finite sample breakdown point is always $1/n$. This suggests that in these cases there is a potentially big scope for improvement and more effort should be put into the development of more robust G -equivariant estimators.

Example 2.8. Let \mathcal{G} be the graph $\overset{2}{\bullet} - \bullet - \overset{3}{\bullet}$. Here $Q = 2$ and hence

$$\Delta(P) = \sum \{P(B) : B \text{ is a linear subspace of dimension } \leq 1\}.$$

In other words this is equal to k/n where k is the maximal number of observations which lie on a single line through the origin. If $\mathbf{x}_{(n)}$ is generic then $k = 1$ and by (9) $\text{fsbp}(T_n, \mathbf{x}_{(n)}, D) \leq \frac{1}{n} \lfloor \frac{n}{2} \rfloor$. Of course for special data sets this is not necessarily true.

Let \mathcal{G} be the four-cycle then $Q = 1$ and hence

$$\Delta(P) = \sum \{P(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m\}.$$

For the finite sample case this is equal to k/n where k is the maximal number of counts of a particular observation in $\mathbf{x}_{(n)}$. If $\mathbf{x}_{(n)}$ is generic then $k = 1$ and hence in this case $\text{fsbp}(T_n, \mathbf{x}_{(n)}, D) \leq \frac{1}{n} \lfloor \frac{n+1}{2} \rfloor$ for any G -equivariant estimator T_n .

In the following section we analyze the symmetry group of Gaussian graphical models and the resulting bounds on breakdown points.

3. THE CONTINUOUS PART OF G

3.1. The clique poset. In this section we discuss various combinatorial constructions related to an undirected graph. They will all be important in our analysis of the group G . We first introduce some basic concepts of the theory of partially ordered sets. For further details see for example [21]. A *partially ordered set* (or *poset*) $P = (P, \leq)$ is a set together with a binary relation satisfying:

1. For all $x \in P$, $x \preccurlyeq x$.
2. If $x \preccurlyeq y$ and $y \preccurlyeq x$, then $x = y$.

3. If $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x \preccurlyeq z$.

If \preccurlyeq satisfies 1. and 3. we say that it defines a *preorder*. We write $x \succcurlyeq y$ for $y \preccurlyeq x$ and $x \prec y$ for $x \preccurlyeq y$ but $y \not\preccurlyeq x$. If \preccurlyeq is a partial order then we say that y *covers* x (or x *is covered by* y) if $x \prec y$ and in addition $x \preccurlyeq z \preccurlyeq y$ implies that either $z = y$ or $x = z$. This concept leads to definition of the *Hasse diagram* of P which is a directed acyclic graph with the set of nodes given by P and the edges indicating the covering relation. The Hasse diagram is depicted so that the smallest elements are always below the larger ones.

Let \mathcal{G} be a graph with the set of vertices given by $[m]$ and the set of edges E . We say that a subset $C \subseteq [m]$ is a *clique (in \mathcal{G})* if every two nodes in C are connected by an edge. We say that C is a *maximal clique* if C is a clique and is maximal (with respect to inclusion) with such a property. We denote the set of maximal cliques by \mathcal{C} . In addition, for every $i \in [m]$ we define \mathcal{C}_i to be the subset of \mathcal{C} containing i , that is

$$\mathcal{C}_i := \{A \in \mathcal{C} : i \in A\} \subseteq 2^{[m]}.$$

By $\mathbf{I}(\mathcal{C})$ the set of subsets of $[m]$ containing \mathcal{C} and closed under taking intersections. If $\mathcal{C} = \{C_1, \dots, C_k\}$ then elements of $\mathbf{I}(\mathcal{C})$ are of the form $\bigcap_{i \in B} C_i$ for some $B \subseteq \{1, \dots, k\}$. The elements of $\mathbf{I}(\mathcal{C})$ are ordered by inclusion, which gives to $\mathbf{I}(\mathcal{C})$ a poset structure.

Recall the preorder defined in (2). The following result gives us equivalent formulation of the ordering relation.

Lemma 3.1. *The following are equivalent to $i \preccurlyeq j$:*

- (i) $N(j) \cup \{j\} \subseteq N(i) \cup \{i\}$
- (ii) for all $C \in \mathcal{C}$ if $j \in C$ then $i \in C$,
- (iii) for all $C \in \mathbf{I}(\mathcal{C})$ if $j \in C$ then $i \in C$,
- (iv) $\mathcal{C}_j \subseteq \mathcal{C}_i$.

Proof. Condition (i) is just the definition of $i \preccurlyeq j$. The implication (i) \Rightarrow (ii) follows from the fact that

$$N(i) = \{k \in [m] : \exists C \in \mathcal{C} \text{ such that } i, k \in C\}.$$

The implication (ii) \Rightarrow (iii) is immediate. Condition (iv) is just a rewording of (ii). Finally, to show that (iv) implies (i) first note that necessarily $j \in N(i)$. Now let $k \in N(j)$. This implies that there exists $C \in \mathcal{C}$ such that j, k and hence also i lie in C . Hence either $k = i$ or $k \in N(i)$. \square

As it was explained in the introduction, this preorder induces an equivalence relation on $[m]$ by $i \sim j$ if $N(i) \cup \{i\} = N(j) \cup \{j\}$. By $\bar{i} \subseteq [m]$ we denote the equivalence class of $i \in [m]$. Now on $[m]/\sim$ we have an induced partial order.

Definition 3.2. We say $\bar{i} \preccurlyeq \bar{j}$ for $\bar{i}, \bar{j} \in [m]/\sim$ if $i \preccurlyeq j$ for some (and hence any) $i \in \bar{i}$, $j \in \bar{j}$. The resulting partially ordered set (poset) on the equivalence classes is denoted by $\mathbf{P}_{\mathcal{C}}$.

In Figure 1 we give three basic graphs and Hasse diagrams of the corresponding posets $\mathbf{P}_{\mathcal{C}}$. The poset $\mathbf{P}_{\mathcal{C}}$ was first introduced in [15]. We note in passing that not all posets $\mathbf{P}_{\mathcal{C}}$ arise as $\mathbf{P}_{\mathcal{C}}$ for some \mathcal{G} . A simple example is given by two posets in Figure 2. Now, using the developed theory, we analyze formally the group G defined by (1).

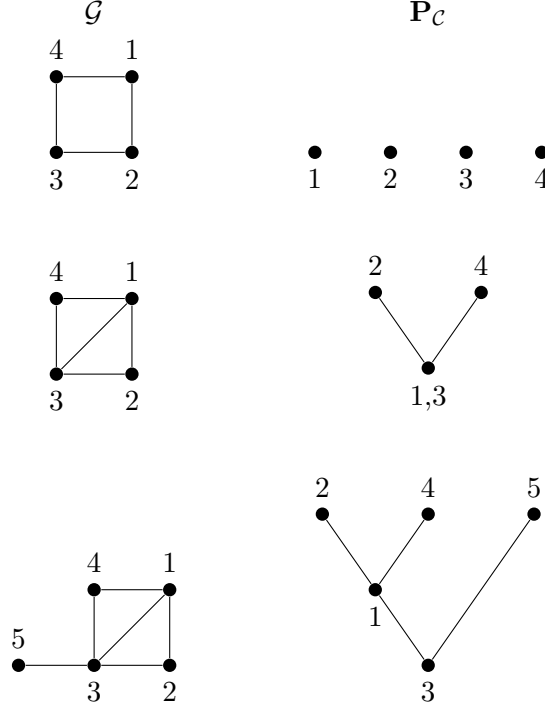


FIGURE 1. Three examples of graphs and corresponding posets $\mathbf{P}_{\mathcal{C}}$ represented by their Hasse diagrams.



FIGURE 2. Two examples of posets that cannot arise as $\mathbf{P}_{\mathcal{C}}$ for some graph with the set of maximal cliques \mathcal{C} .

3.2. The connected component of the identity. Recall the symmetry group G of a Gaussian graphical model defined in the introduction. By (1) this is a closed subgroup of $\mathrm{GL}_m(\mathbb{R})$ and hence a Lie group. Some basic results in the Lie group theory have been provided for reader's convenience in the appendix. Denote by G^0 the connected component of G containing the identity matrix I . By standard Lie group theory, G^0 is uniquely determined by its Lie algebra \mathfrak{g}^0 . We first exploit the fact that G^0 contains the set of all diagonal matrices $T^m := (\mathrm{GL}_1(\mathbb{R}))^m \subseteq \mathrm{GL}_m(\mathbb{R})$. Let $\mathcal{G} = ([m], E)$ be an undirected graph and let $\mathbf{P}_{\mathcal{C}}$ be its associated poset defined in the previous subsection. We obtain

the following result, which also shows that the group G^0 defined here coincides with G^0 defined in the introduction.

Proposition 3.3. *For every $i, j \in [m]$ the element E_{ij} lies in \mathfrak{g}^0 if and only if $j \prec i$. In particular, for every $g \in G^0$ we have $g_{ij} = 0$ whenever $j \not\prec i$.*

Proof. By Lemma A.1 it remains to check which E_{ij} lie in \mathfrak{g}^0 , or equivalently, which $(I + tE_{ij})$ lie in G^0 for all $t \in \mathbb{R}$. We check for which $i \neq j \in [m]$

$$(10) \quad (I + tE_{ji})K(I + tE_{ij}) \in \mathcal{S}_{\mathcal{G}},$$

whenever $K \in \mathcal{S}_{\mathcal{G}}$. We have:

$$(11) \quad (I + tE_{ji})K(I + tE_{ij}) = K + t(E_{ji}K + KE_{ij}) + t^2E_{ji}KE_{ij}.$$

Denote by $e_k \in \mathbb{R}^m$ the unit vector with 1 on the coordinate corresponding to $k \in [m]$. Since for every $k \neq l \in [m]$ necessarily $e_k^T E_{ji} K E_{ij} e_l = 0$, the matrix in (11) lies in $\mathcal{S}_{\mathcal{G}}$ if and only if

$$(12) \quad e_k^T (E_{ji}K + KE_{ij}) e_l = 0 \quad \text{for every } (k, l) \notin E.$$

If we take (k, l) equal to (i, j) then $e_k^T (E_{ji}K + KE_{ij}) e_l = K_{ii}$ and hence (i, j) needs to be an edge of \mathcal{G} . The condition in (12) is automatically satisfied if $j \neq k, l$. Now we check the case $j = l$ (but $i \neq k$). In this case (12) is satisfied only if $K_{ki} = 0$ or equivalently $(i, k) \notin E$. In other words for every $(j, k) \notin E$ also $(k, i) \notin E$, which translates into $N(i) \cup i \subseteq N(j) \cup j$ and consequently to $j \prec i$. If $j = k$ (but $i \neq l$) then we get the same condition. \square

Example 3.4. Consider a simple graph $A_3 : \overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$. By Proposition 3.3 the Lie algebra \mathfrak{g}^0 is generated by E_{11}, E_{22}, E_{33} together with E_{21} and E_{31} . The element E_{21} lies in \mathfrak{g}^0 because $N(2) \cup 2 = \{1, 2\} \subseteq N(1) \cup 1 = \{1, 2, 3\}$. This also shows that E_{12} does not lie in \mathfrak{g}^0 . The group G^0 consists of matrices of the form

$$\begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & 0 & * \end{bmatrix},$$

where the asterisk denotes an element which can be non-zero.

Remark 3.5. The group G^0 has been studied for other Gaussian models. For conditional independence lattice models this group was named the group of generalized block-triangular matrices with lattice structure (see [2, Section 2.4]). The link between the conditional independence lattice models and certain Gaussian graphical models is discussed in [1].

4. THE FULL STRUCTURE OF G

In general G may contain elements which are not in G^0 . For example for the graph A_3 the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

represents the only non-trivial automorphism of this graph, lies also in G but it does not lie in G^0 . Recall that a graph automorphism σ of $\mathcal{G} = ([m], E)$ is a bijection of $[m]$ such that $(u, v) \in E$ if and only if $(\sigma(u), \sigma(v)) \in E$. Any such automorphism gives rise to an element of G by taking the corresponding permutation matrix and thus $\text{Aut}(\mathcal{G}) \subseteq G$.

Recall from the introduction the definition of the induced graph $\tilde{\mathcal{G}}$ on the set of equivalence classes $[m]/\sim$. Note that we have a (non-canonical) embedding $\text{Aut}(\tilde{\mathcal{G}}, c) \rightarrow \text{Aut}(\mathcal{G}) \subseteq G$ defined as follows. In each equivalence class of \tilde{i} choose an arbitrary labelling of the elements. Then, every automorphism of $(\tilde{\mathcal{G}}, c)$ lifts to a unique automorphism of \mathcal{G} preserving the labelling.

It is a standard result from the Lie group theory that the connected component of the identity G^0 is a normal subgroup of G . Hence, to show that $G = G^0 \rtimes \text{Aut}(\tilde{\mathcal{G}}, c)$, in order to prove Theorem 1.1, we need to show that $G = G^0 \cdot \text{Aut}(\tilde{\mathcal{G}}, c)$ and $G^0 \cap \text{Aut}(\tilde{\mathcal{G}}, c) = \{I\}$. For this, it suffices to construct a surjective homomorphism

$$(13) \quad \sigma : G \longrightarrow \text{Aut}(\tilde{\mathcal{G}}, c), \quad g \mapsto \sigma_g$$

such that $\ker(\sigma) = G^0$. To construct the homomorphism (13), consider now the family

$$F := \{a \in \mathbb{R}^m : aa^T \in \mathcal{S}_{\mathcal{G}}\}.$$

For $a \in F$ and $g \in G$ we have $g^T aa^T g \in \mathcal{S}_{\mathcal{G}}$ so that $g^T a \in F$. Necessary and sufficient for $a \in F$ is that $\text{supp}(a)$ defined by

$$\text{supp}(a) := \{i \in [m] : a_i \neq 0\}$$

is a clique in \mathcal{G} . In particular $e_i \in F$ for all $i \in [m]$. Hence $\text{supp}(g^T e_i)$ is a clique, which we denote by $R_g(i)$.

We use R_g also to denote the induced map $R_g : 2^{[m]} \rightarrow 2^{[m]}$:

$$(14) \quad R_g(A) = \bigcup_{i \in A} R_g(i).$$

Note that directly from the definition it follows that

$$(15) \quad R_g\left(\bigcup_{i=1}^l A_i\right) = \bigcup_{i=1}^l R_g(A_i)$$

for any subsets $A_1, \dots, A_l \subseteq [m]$. Denote by \mathcal{O}_A the $|A|$ -dimensional linear subspace of \mathbb{R}^m of all points $x = (x_1, \dots, x_m)$ with $x_i = 0$ for all $i \in [m] \setminus A$. Since g is invertible we conclude that

$$(16) \quad |R_g(A)| \geq |A| \quad \text{for every subset } A \subseteq [m].$$

This follows from the fact that g^T maps the subspace \mathcal{O}_A of \mathbb{R}^m into $\mathcal{O}_{R_g(A)}$ so the dimension of $\mathcal{O}_{R_g(A)}$ has to be at least the dimension of \mathcal{O}_A .

Lemma 4.1. *For every $g \in G$ if C is a clique of \mathcal{G} , the set $R_g(C)$ is a clique as well. Moreover, if C is a maximal clique then $R_g(C)$ is maximal as well, and of the same cardinality.*

Proof. For a generic $a \in F$ with support C the support of $g^T a$ is $R_g(C)$. Hence this must be a clique. If C is a maximal clique then let $D \supseteq R_g(C)$ be a maximal clique containing $R_g(C)$. Since $g^{-T} g^T e_i = e_i$ for every $i \in C$ then, for some $j \in R_g(C)$, $\{i\} \subseteq R_{g^{-1}}(j)$. In particular $C' = R_{g^{-1}}(D)$ is a clique of \mathcal{G} containing C and hence necessarily $C = C'$. Similarly, for every $i \in D$, for some $j \in C$, $\{i\} \subseteq R_g(j)$ and hence $R_g(C) = D$. \square

Denote by $\text{Aut}(\mathbf{I}(\mathcal{C}), c)$ the subgroup of $\text{Aut}(\mathbf{I}(\mathcal{C}))$ preserving cardinality of elements of $\mathbf{I}(\mathcal{C})$. We have the following important result.

Proposition 4.2. *For every $g \in G$ the map $M \mapsto R_g(M)$ defines an automorphism $\rho_g : \mathbf{I}(\mathcal{C}) \rightarrow \mathbf{I}(\mathcal{C})$. The map $\rho : G \rightarrow \text{Aut}(\mathbf{I}(\mathcal{C}), c)$ given by $g \mapsto \rho_g$ is an antihomomorphism. Moreover, $\ker \rho = G^0$.*

Proof. First we have to show that ρ_g is a well defined order preserving automorphism of $\mathbf{I}(\mathcal{C})$ for every $g \in G$. Let $\mathcal{C} = \{C_1, \dots, C_k\}$ then

$$M = \bigcap_{i \in A} C_i \quad \text{for some } A \subseteq [k].$$

Since M is a clique, $R_g(M)$ is a clique as well by Lemma 4.1. Also, by (14), if $k \in R_g(M) = \bigcup_{j \in M} R_g(j)$ then there exists $l \in M$ such that $k \in R_g(l)$ and $R_g(l) \subseteq R_g(C_i)$ for all $i \in A$. It follows that $R_g(M) \subseteq \bigcap_{i \in A} R_g(C_i)$. Hence

$$(17) \quad |M| \leq |R_g(M)| \leq \left| \bigcap_{i \in A} R_g(C_i) \right|,$$

where the first inequality follows from (16). We use now the same argument for $R_{g^{-1}}$. Since C_i is a maximal clique, $R_{g^{-1}}(R_g(C_i))$ is a maximal clique. Moreover, $g^{-T} g^T e_i = e_i$ for every $i \in [m]$ and hence, for some $j \in R_g(C_i)$, $i \in R_{g^{-1}}(j)$, which implies that

$$R_{g^{-1}} R_g(C_i) = C_i$$

by maximality of both cliques. It follows that

$$R_{g^{-1}} \left(\bigcap_{i \in A} R_g(C_i) \right) \subseteq \bigcap_{i \in A} R_{g^{-1}} R_g(C_i) = M$$

and consequently

$$(18) \quad \left| \bigcap_{i \in A} R_g(C_i) \right| \leq |R_{g^{-1}} \left(\bigcap_{i \in A} R_g(C_i) \right)| \leq |M|,$$

where again the first inequality follows from (16). By (17) and (18) we have that

$$|M| = |R_g(M)| = \left| \bigcap_{i \in A} R_g(C_i) \right|.$$

This shows that R_g preserves cardinality of elements of $\mathbf{I}(\mathcal{C})$. Moreover, together with the fact that $R_g(M) \subseteq \bigcap_{i \in A} R_g(C_i)$ this implies that

$$(19) \quad R_g \left(\bigcap_{i \in A} C_i \right) = \bigcap_{i \in A} R_g(C_i).$$

This proves that ρ_g is a set automorphism of $\mathbf{I}(\mathcal{C})$ preserving cardinality. To show that ρ_g preserves the order let $M \subset N$ be two elements in $\mathbf{I}(\mathcal{C})$. Then $M = \bigcap_{i \in A} C_i$ and $N = \bigcap_{i \in B} C_i$ for $B \subset A$. By (19) also $R_g(M) \subseteq R_g(N)$, which proves that $\rho_g \in \text{Aut}(\mathbf{I}(\mathcal{C}), c)$ for every $g \in G$.

To show that $\rho : \mathcal{G} \rightarrow \text{Aut}(\mathbf{I}(\mathcal{C}), c)$ defines a group antihomomorphism, it suffices to show that $R_g R_h(C) = R_{hg}(C)$ for every $g, h \in G$ and $C \in \mathcal{C}$. If $j \in R_h(i)$ for $i \in C$ and $k \in R_g(j)$ then $k \in \text{supp}(g^T h^T e_i) = \text{supp}((hg)^T e_i)$ so $k \in R_{hg}(i)$ and hence $R_g R_h(C) \subseteq R_{hg}(C)$. Since both sets are maximal cliques by Lemma 4.1 they must be equal.

We now show that $\ker(\rho) = G^0$. If $g \in G^0$ then g_{ij} can be nonzero (or $j \in R_g(i)$) only if $j \preceq i$. If C is a maximal clique then $j \in R_g(C)$ only if there exists $i \in C$ such that $j \preceq i$. In particular $j \in C$ as well because $\mathcal{C}_i \subseteq \mathcal{C}_j$, and hence $R_g(C) = C$ (because $R_g(C) \subseteq C$ and is maximal). This shows that if $g \in G^0$ then ρ_g is the identity, so $G^0 \subseteq \ker(\rho)$.

Let now ρ_g be the identity of $\mathbf{I}(\mathcal{C})$. Take any $i, j \in [m]$ such that $j \not\preceq i$ or equivalently, by Lemma 3.1, $\mathcal{C}_i \not\subseteq \mathcal{C}_j$. Take $C \in \mathcal{C}_i \setminus \mathcal{C}_j$. Since $R_g(C) = C$ then in particular $j \notin R_g(i)$ or equivalently $g_{ij} = 0$, which implies that $g \in G^0$. This shows that also $\ker(\rho) \subseteq G^0$. \square

In the remaining part of this section we study the connection between $\text{Aut}(\mathbf{I}(\mathcal{C}), c)$ and $\text{Aut}(\tilde{\mathcal{G}}, c)$. Since, by Lemma 3.1, $j \preceq i$ if and only if $\mathcal{C}_i \subseteq \mathcal{C}_j$ then

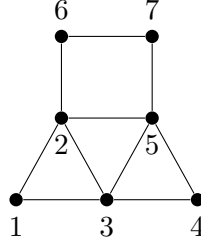
$$(20) \quad \downarrow i \quad := \quad \{j \in [m] : j \preceq i\} \quad = \quad \bigcap_{C \in \mathcal{C}_i} C$$

and hence $\downarrow i$ lies in $\mathbf{I}(\mathcal{C})$. We say that $A \in \mathbf{I}(\mathcal{C})$ is *join-irreducible* if $A \neq \emptyset$ and A cannot be expressed as $B \cup C$ for $B, C \in \mathbf{I}(\mathcal{C})$ such that $B, C \subset A$. The set of all join-irreducible elements of $\mathbf{I}(\mathcal{C})$ is denoted by $\mathbf{J}(\mathcal{C})$. The elements in $\mathbf{I}(\mathcal{C})$, which are smaller than $\downarrow i$, are obtained by intersecting $\downarrow i$ with some $C \notin \mathcal{C}_i$. Since, for every $C \notin \mathcal{C}_i$ we have $i \notin (\downarrow i) \cap C$ then $\downarrow i \in \mathbf{J}(\mathcal{C})$. On the other hand, if $A \in \mathbf{J}(\mathcal{C})$ then there exists $i \in A$ such that $i \notin C$ for every $C \subset A$ in $\mathbf{I}(\mathcal{C})$. Suppose that there exists $D \in \mathcal{C}_i$ such that $A \not\subseteq D$. Then $C = A \cap D \subset A$ and $i \in C$ which leads to a contradiction and hence if $D \in \mathcal{C}_i$ then $A \subseteq D$ or equivalently $A \subseteq \bigcap_{C \in \mathcal{C}_i} C$ in $\mathbf{I}(\mathcal{C})$. Since $i \in A$ then also $A = \bigcap_{C \in \mathcal{C}_i} C$ and consequently every join-irreducible element of $\mathbf{I}(\mathcal{C})$ is of the form $\downarrow i$.

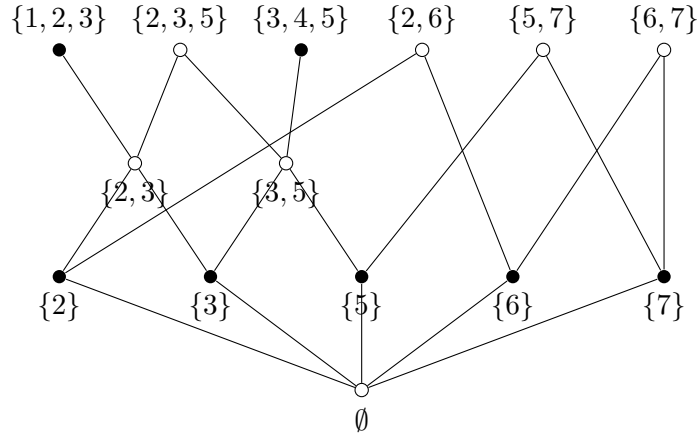
Lemma 4.3. *The map $\phi : \bar{i} \mapsto \downarrow i$ is an order isomorphism $\mathbf{P}_{\mathcal{C}} \rightarrow \mathbf{J}(\mathcal{C})$.*

Proof. Since $\downarrow i$ is join-irreducible and it does not depend on the representative of \bar{i} , ϕ is well defined. Since every $A \in \mathbf{J}(\mathcal{C})$ is of the form $\downarrow i$ then ϕ is surjective. It is also injective and hence it is a bijection. It preserves the ordering because $i \preceq j$ means $\mathcal{C}_j \subseteq \mathcal{C}_i$ and hence $\downarrow i \subseteq \downarrow j$ by (20). \square

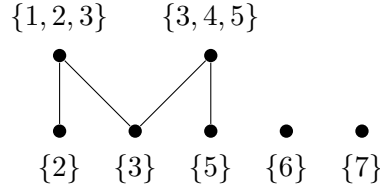
Example 4.4. Consider the following graph:



Then there are six maximal cliques given by $C_1 = \{1, 2, 3\}$, $C_2 = \{2, 3, 5\}$, $C_3 = \{3, 4, 5\}$, $C_4 = \{2, 6\}$, $C_5 = \{5, 7\}$ and $C_6 = \{6, 7\}$. The poset $\mathbf{I}(\mathcal{C})$ is given by:



where the maximal element corresponds to an empty intersection and the minimal to the intersection of all maximal cliques. The join irreducible elements of $\mathbf{I}(\mathcal{C})$ are depicted as solid nodes. Hence the poset $\mathbf{J}(\mathcal{C})$ is given by:



Now we are ready to prove the first main theorem of this paper.

Proof of Theorem 1.1. We define a homomorphism $G \rightarrow \text{Aut}(\widetilde{G}, c)$ as follows. Every $g \in G$ defines an automorphism ρ_g of $\mathbf{I}(\mathcal{C})$ preserving cardinality as defined in Proposition 4.2. This automorphism ρ_g restricts to an automorphism of $\mathbf{J}(\mathcal{C})$ because $\mathbf{J}(\mathcal{C})$ consists of all the non-empty elements of $\mathbf{I}(\mathcal{C})$ that cover at most one element in $\mathbf{I}(\mathcal{C})$ and this property has to be preserved by every poset automorphism of $\mathbf{I}(\mathcal{C})$. Therefore, by Lemma 4.3, ρ_g induces a bijection on $[m]/\sim$, which we denote by σ_g . More concretely σ_g is uniquely defined by

$$\rho_g(\downarrow i) = \downarrow \sigma_g(\bar{i}) \quad \text{for every } \bar{i} \in [m]/\sim.$$

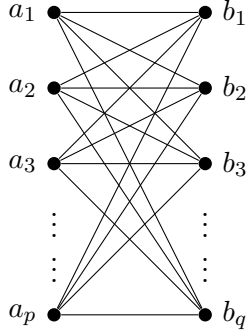


FIGURE 3. The complete bipartite graph.

Since by (20) we have $\downarrow \sigma_g(i) = \bigcap_{C \in \mathcal{C}_{\sigma_g(i)}} C$ and by (19) $\rho_g(\downarrow i) = \bigcap_{C \in \mathcal{C}_i} \rho_g(C)$ then it follows that

$$\mathcal{C}_{\sigma_g(\bar{i})} = \rho_g(\mathcal{C}_i) := \{\rho_g(C) : C \in \mathcal{C}_i\}.$$

Two equivalence classes \bar{i}, \bar{j} are linked in \tilde{G} if and only if there exists $C \in \mathcal{C}$ such that $\bar{i}, \bar{j} \in C$ or equivalently $\mathcal{C}_i \cap \mathcal{C}_j \neq \emptyset$. In this case also $\mathcal{C}_{\sigma_g(\bar{i})} \cap \mathcal{C}_{\sigma_g(\bar{j})} \neq \emptyset$, which shows that σ_g induces a graph automorphism of \tilde{G} . The fact that this automorphism will always preserve the cardinality of equivalence classes follows from the fact that ρ_g preserves cardinality of elements of $\mathbf{I}(\mathcal{C})$. Denote by σ^{-1} the homomorphism $G \rightarrow \text{Aut}(\tilde{G}, c)$ defined by $g \mapsto \sigma_g^{-1}$. Since $\ker \sigma = \ker \rho$ then by Proposition 4.2 $\ker \sigma = G^0$. \square

5. SPECIAL GRAPHS AND SMALL EXAMPLES

Assume that a graph \mathcal{G} with set of nodes given by $[m]$ is given in form of the set \mathcal{C} of its maximal cliques. The computation of G^0 can be done efficiently by listing all \mathcal{C}_i for $i \in [m]$. Then by Lemma 3.1 (iv) we have $i \preccurlyeq j$ if and only if $\mathcal{C}_j \subseteq \mathcal{C}_i$. On the other hand, if \mathcal{G} is given explicitly in terms of the edge set E it is more efficient to use Lemma 3.1 (i) and for each $i \in [m]$ to compute $N(i) \cup i = \{j \in [m] : (i, j) \in E\} \cup i$. The computation of the discrete part $\text{Aut}(\tilde{\mathcal{G}}, c)$ is in general much harder and it depends on current algorithms for computing automorphism groups of graphs.

Let S_m denote the symmetric group on $[m]$, D_m the dihedral group of graph isomorphisms of an m -cycle. Also recall that $T^k \simeq (\text{GL}_1(\mathbb{R}))^k$ denotes the group of all diagonal invertible $k \times k$ matrices. In Table 1 we provide the full description of G for all undirected graphs on $m = 2, 3, 4$ vertices.

In the rest of this section we also compute G for some interesting families of graphs. Consider for example the complete bipartite graph in Figure 3. The maximal cliques containing a_i are $\{\{a_i, b_j\} : j = 1, \dots, q\}$ for every $i = 1, \dots, p$ and the maximal cliques containing b_j are $\{\{a_i, b_j\} : i = 1, \dots, p\}$ for every $j = 1, \dots, q$. If $p = q = 1$ then $G = \text{GL}_2(\mathbb{R})$. If $p = 1, q \geq 2$ then the poset $\mathbf{P}_{\mathcal{C}}$ looks as follows.













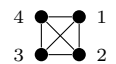

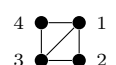

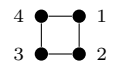
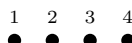
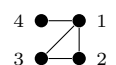

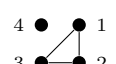

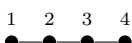

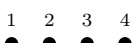

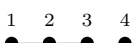
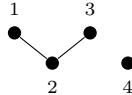
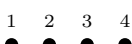

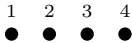
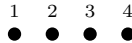
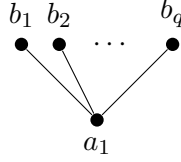
| \mathcal{G} | \mathbf{P}_c | G^0 | $\text{Aut}(\tilde{\mathcal{G}}, c)$ |
|---|---|--|--------------------------------------|
|  |  | $\text{GL}_2(\mathbb{R})$ | $\{\text{id}\}$ |
|  |  | T^2 | S_2 |
|  |  | $\text{GL}_3(\mathbb{R})$ | $\{\text{id}\}$ |
|  |  | $\begin{bmatrix} * & * & 0 \\ 0 & * & 0 \\ 0 & * & * \end{bmatrix}$ | S_2 |
|  |  | $\text{GL}_2(\mathbb{R}) \times T^1$ | $\{\text{id}\}$ |
|  |  | T^3 | S_3 |
|  |  | $\text{GL}_4(\mathbb{R})$ | $\{\text{id}\}$ |
|  |  | $\begin{bmatrix} * & 0 & * & 0 \\ * & * & * & 0 \\ * & 0 & * & 0 \\ * & 0 & * & * \end{bmatrix}$ | S_2 |
|  |  | T^4 | D_4 |
|  |  | $\begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ * & 0 & 0 & * \end{bmatrix}$ | $\{\text{id}\}$ |
|  |  | $\text{GL}_3(\mathbb{R}) \times T^1$ | $\{\text{id}\}$ |
|  |  | $\begin{bmatrix} * & 0 & 0 & 0 \\ * & * & 0 & 0 \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}$ | S_2 |
|  |  | $\text{GL}_2(\mathbb{R}) \times \text{GL}_2(\mathbb{R})$ | S_2 |
|  |  | $\begin{bmatrix} * & * & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & * & * & 0 \\ 0 & 0 & 0 & * \end{bmatrix}$ | S_2 |
|  |  | $\text{GL}_2(\mathbb{R}) \times T^1 \times T^1$ | S_2 |
|  |  | T^4 | S_4 |

TABLE 1. Small undirected graphs \mathcal{G} , corresponding groups G^0 and $\text{Aut}(\tilde{\mathcal{G}}, c)$ up to isomorphism.



Finally, if $p, q \geq 2$ then none of the nodes are comparable. It follows that $G^0 = T^{p+q}$. Since $\mathcal{G} = \tilde{\mathcal{G}}$ and $\text{Aut}(\tilde{\mathcal{G}}, c) = S_p \times S_q$, we obtain the following result.

Proposition 5.1. *Let \mathcal{G} be the bipartite graph in Figure 3 with $p, q \geq 2$. Then*

$$G = T^{p+q} \rtimes (S_p \times S_q).$$

If \mathcal{G} is an m -cycle then if $m \geq 4$ we have that $\mathcal{C} = \{\{1, 2\}, \dots, \{m-1, m\}, \{m, 1\}\}$ and, as above, the poset $\mathbf{P}_{\mathcal{C}}$ is trivial with none of the nodes comparable. In this case $\text{Aut}(\tilde{\mathcal{G}}, c)$ is the dihedral group D_m . If \mathcal{G} has no edges then $\text{Aut}(\tilde{\mathcal{G}}, c) = S_m$. Finally, if \mathcal{G} is the complete graph on m nodes then $\tilde{\mathcal{G}}$ is a simple node with cardinality m and hence there are no non-trivial automorphisms of $\tilde{\mathcal{G}}$. We obtain the following result.

Proposition 5.2. *Let \mathcal{G} be an undirected graph with m vertices. Then*

- if \mathcal{G} is the m -cycle for $m \geq 4$.

$$G = T^m \rtimes D_m.$$

- if \mathcal{G} is a graph with $m \geq 2$ vertices and no edges

$$G = T^m \rtimes S_m.$$

- if \mathcal{G} is the complete graph

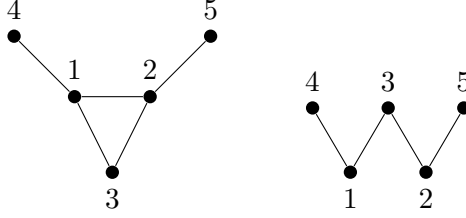
$$G = \text{GL}_m(\mathbb{R}).$$

We conclude this section with a special small graph - the *bull graph* - which is a graph on five vertices depicted in Figure 4. The terminology is taken from the graph theory (see for example [6]). This simple graph is very special because it is the smallest graph such that there exists an element in $\mathbf{P}_{\mathcal{C}}$ covering more than one element, which is relevant in the study of G -orbits in $\mathcal{S}_{\mathcal{G}}^+$. We say that a graph \mathcal{G} is *bull-free* it does not contain a bull graph as an induced subgraph. We have the following result.

Lemma 5.3. *Let \mathcal{G} be a bull-free undirected graph. Then in $\mathbf{P}_{\mathcal{C}}$ every element covers at most one element.*

Proof. Suppose that k covers i and j in $\mathbf{P}_{\mathcal{C}}$. This implies that $N(k) \cup k \subseteq N(i) \cup i$ and $N(k) \cup k \subseteq N(j) \cup j$, where neither $N(i) \cup i \subseteq N(j) \cup j$ nor $N(j) \cup j \subseteq N(i) \cup i$. It follows that we can find vertices l and m such that $l \in N(i) \setminus (N(k) \cup k)$ and $l \in N(j) \setminus (N(k) \cup k)$ such that the induced subgraph on i, j, k, l, m is the bull graph. \square

The existence of a bull graph as an induced subgraph in \mathcal{G} does not imply that we can find an element in $\mathbf{P}_{\mathcal{C}}$ covering more than one element. A simple example is obtained by adding to the bull graph in Figure 4 an additional vertex and joining it with the vertex 3.

FIGURE 4. The bull graph and the corresponding \mathbf{P}_C on the right.

6. G -ORBITS IN \mathbb{R}^m AND G -EQUIVARIANCE

The final aim of this section is to prove Theorem 1.2 and Theorem 2.6. Before we do this we need to formally analyze G -stabilizers of points in \mathbb{R}^m , or more specifically, pointwise stabilizers of subsets $B \subseteq \mathbb{R}^m$. By duality between G -stabilizers and G -orbits we proceed with the analysis of G -orbits of \mathbb{R}^m in Section 6.1. To complete our discussion in Section 6.2 we provide the conditions on the sample size for the existence of a G -equivariant estimator.

6.1. G -orbits in the sample space. Since G contains the group of all invertible diagonal matrices T^m , every vector $\mathbf{x} = (x_1, \dots, x_m)$ with support $A = \{i : x_i \neq 0\} \subseteq [m]$ lies in the same orbit as the vector $\text{ind}(\mathbf{x}) \in \{0, 1\}^m$ with the same support. Hence all that matters in the identification of G -orbits is the support of a vector. Recall from Section 4 that for $A \subseteq [m]$ by \mathcal{O}_A we denote the set of all points in \mathbb{R}^m with the support equal to A .

The following lemma shows that, when searching for the supremum in (6), we can restrict to elements $g \in G^0 \cap G_1$. This shows that for our analysis it is sufficient to understand G^0 -orbits, which will be the main focus of this section.

Lemma 6.1. *If the pointwise stabiliser in G of a subset $B \in \mathcal{B}(\mathbb{R}^m)$ contains an element g such that $\det g \neq \pm 1$, then it contains an element of $G^0 \setminus \{I\}$.*

Proof. Let g be a pointwise stabiliser of B such that $\det g \neq \pm 1$. Then also g^l is a pointwise stabiliser of B . By Theorem 1.1 every $g \in G$ can be written as $g = g'\sigma$, where σ is a permutation matrix and $g' \in G^0$. Since $\text{Aut}(\tilde{\mathcal{G}}, c)$ is a finite group then there exists $l_0 \geq 1$ be such that $\sigma^{l_0} = I$. In this case $g^{l_0} = g'^{l_0} \in G^0$. Because $\det g \neq \pm 1$, $g' \neq I$, and hence $g' \in G^0 \setminus \{I\}$. \square

Let P be a poset and $Q \subset P$. We say that Q is a *down-set* if, whenever $x \in Q$, $y \in P$ and $y \leq x$, we have $y \in Q$. Dually, Q is an *up-set* if, whenever $x \in Q$, $y \in P$ and $y \geq x$, we have $y \in Q$. The family of all down-sets of P is denoted by $\mathbf{O}^\downarrow(P)$ and the family of all up-sets of P by $\mathbf{O}^\uparrow(P)$. They themselves are posets, under the inclusion order. For any $A \subseteq P$ define the down-set associated to

$$\downarrow A := \{x \in P : \exists a \in A \text{ such that } x \leq a\}.$$

We write $\downarrow a$ for $\downarrow \{a\}$, which is consistent with the notation introduced in (20). Dually we define $\uparrow A$ and $\uparrow a$.

Let $\mathfrak{o} : 2^{[m]} \rightarrow \mathbf{O}^\uparrow(\mathbf{P}_C)$ be defined by $A \mapsto \uparrow A$. This function defines an equivalence relation on $2^{[m]}$ by $A \sim B$ if $\uparrow A = \uparrow B$ in $\mathbf{O}^\uparrow(\mathbf{P}_C)$. If we identify each $A \in 2^{[m]}$ with the open subset \mathcal{O}_A then, as the following result shows, the \mathfrak{o} -equivalence classes correspond to G^0 -orbits.

Proposition 6.2. *The G^0 -orbits of \mathbb{R}^m are in one-to-one correspondence with the up-sets of \mathbf{P}_C or equivalently with \mathfrak{o} -congruence classes in $2^{[m]}$. The orbits are*

$$\overline{\mathcal{O}}_A := \bigcup_{\uparrow B = A} \mathcal{O}_B \quad \text{for all } A \in \mathbf{O}^\uparrow(\mathbf{P}_C).$$

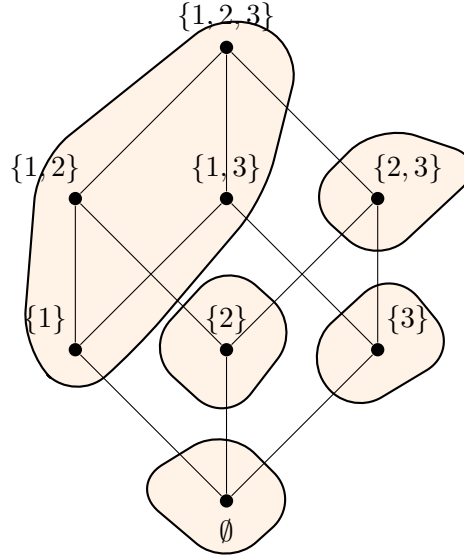
In particular $\dim(\overline{\mathcal{O}}_A) = |A|$ and hence the unique dense orbit is given by $\overline{\mathcal{O}}_{[m]}$.

Proof. If $i \succ j$ the element $I + tE_{ij}$ lies in G^0 . Since $(I + tE_{ij})\mathbf{x}$ differs from \mathbf{x} only on the i -th coordinate where it is equal to $x_i + tx_j$ we conclude that:

- (a) if $i, j \in A$ then $I + tE_{ij}$ stabilizes \mathcal{O}_A ;
- (b) if $j \in A, i \notin A$ then $I + tE_{ij}$ merges \mathcal{O}_A and $\mathcal{O}_{A \cup i}$;
- (c) if $j \notin A$ then $I + tE_{ij}$ stabilizes \mathcal{O}_A .

Hence the orbits of G^0 correspond to subsets $A \subseteq [m]$ closed under the above operations. Namely, if $j \in A$ then $i \in A$ for all $i \succ j$. These are precisely the up-sets of \mathbf{P}_C . \square

Example 6.3. Let \mathcal{G} be the graph $\overset{2}{\bullet} - \overset{1}{\bullet} - \overset{3}{\bullet}$. In this case we have 5 orbits of G^0 represented by $(0, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(0, 1, 1)$, $(1, 0, 0)$. This corresponds to all congruence classes in the Boolean lattice of subsets of $\{1, 2, 3\}$, which are marked on the corresponding Hasse diagram.



The congruence class containing the set $\{1, 2, 3\}$ corresponds to the unique three dimensional dense orbit containing all points with non-zero coordinates. For example, points

$(1, 0, 0)$ and $(1, 1, 1)$ are in the same G -orbit because

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Finally, to understand the G -orbits it suffices to check which G^0 -orbits $\overline{\mathcal{O}}_A$ for $A \in \mathbf{O}^\uparrow(\mathbf{P}_{\mathcal{C}})$ get glued by $\text{Aut}(\tilde{\mathcal{G}}, c)$. In our simple example $\overline{\mathcal{O}}_1$ gets glued with $\overline{\mathcal{O}}_3$. Other orbits are stabilized by $\text{Aut}(\tilde{\mathcal{G}}, c)$. This gives us four distinct G -orbits in \mathbb{R}^3 .

6.2. G -equivariant maps. Let \mathcal{G} be an undirected graph with set of maximal cliques \mathcal{C} . In this section, by proving Theorem 1.2, we find the conditions which assure that there exists a G -equivariant estimator $T_n : (\mathbb{R}^m)^n \rightarrow \mathcal{S}_{\mathcal{G}}^+$. This question is related to the question of existence of the MLE for Gaussian graphical models. If the MLE exists with probability 1 then automatically we have a G -equivariant map T_n . It turns out however that, if \mathcal{G} is not the complete graph K_m , there may be cases when the MLE does not exist but still there are other G -equivariant maps. This will follow from Theorem 1.2.

To present known results on the existence of the MLE we first define decomposable graphs.

Definition 6.4. A undirected graph \mathcal{G} is decomposable if it contains no induced n -cycles for $n \geq 4$.

We define a *decomposable cover* $\mathcal{G}^* = ([m], E^*)$ of an undirected graph $\mathcal{G} = ([m], E)$ as a decomposable graph satisfying $E \subseteq E^*$. The maximal clique size of \mathcal{G} is $q = \max_{C \in \mathcal{C}} |C|$. If \mathcal{C}^* is the set of maximal cliques of \mathcal{G}^* then we also define $q^* = \max_{C \in \mathcal{C}^*} |C|$. Let $\mathbf{x}_{(n)}$ be a random sample of size n . Then, as it is discussed in [22, Corollary 2.3], the MLE will never exists if $n < q$ and will always exists if $n \geq q^*$. Of course this latter bound is optimal when a decomposable cover \mathcal{G}^* is chosen with minimal q^* . This in particular gives the complete characterization in the decomposable case, where we can take $\mathcal{G}^* = \mathcal{G}$. For non-decomposable graphs there is no known closed-form condition if $q \leq n < q^*$ and only special cases were studied (see [22] and references therein).

As indicated in Theorem 1.2 the characterisation of the existence of G -equivariant estimators has an elegant closed form. We now prove this result.

Definition 6.5. We say that some property holds for *generic* k -tuples $(x_1, \dots, x_k) \in (\mathbb{R}^m)^k$ if it holds for (x_1, \dots, x_k) outside the zero set of some non-zero polynomial.

Note that if a property holds for generic k -tuples, then it holds with probability one for the random sample $\mathbf{x}_{(n)}$ drawn from any non-degenerate continuous probability distribution.

Theorem 6.6. *The minimal number k for which the pointwise stabiliser in G^0 of generic k -tuple $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^m$ consists entirely of determinant-one matrices equals $\max_{i \in [m]} |\downarrow i|$. For that value of k the stabiliser of k generic vectors is, in fact, the trivial group $\{I\}$.*

Proof. The condition that $g \in G^0$ fixes a single vector $\mathbf{x} \in \mathbb{R}^m$ translates into m linear conditions on the entries of g , namely:

$$\sum_{j \preccurlyeq i} g_{ij} x_j = x_i \quad \text{for } i = 1, \dots, m.$$

The i -th condition concerns only the entries in the i -th row of g . We therefore concentrate on that single row of g , and consider the entries g_{ij} , $i \succcurlyeq j$ as variables to be solved from the linear equations above as \mathbf{x} ranges through the given k -tuple of vectors. Since the given k -tuple is generic, those equations are linearly independent as long as k is at most the cardinality of $\downarrow i$. Hence they determine the i -th row uniquely as soon as k is at least that number. Hence as soon as k is at least the maximal cardinality of the sets $\downarrow i$ over all i the pointwise stabiliser of the k vectors will be trivial.

What remains to be checked is that for smaller k the pointwise stabiliser of generic $\mathbf{x}_1, \dots, \mathbf{x}_k$ does *not* consist entirely of determinant-1 matrices. This is easiest seen by considering the Lie algebra of that pointwise stabiliser, which is the set of matrices A in the Lie algebra of G satisfying the linear conditions $A\mathbf{x}_1 = \dots = A\mathbf{x}_k = 0$. Let i be a row index for which $\downarrow i$ has more than k elements. Then the linear conditions on A do not fix the i -th row of A uniquely. Moreover, by genericity, they do not even fix the diagonal entry A_{ii} uniquely. As a consequence, they do not determine the trace of A uniquely. This shows that the Lie algebra of the pointwise stabiliser is not contained in the Lie algebra of trace-zero matrices (see Appendix A). But then the pointwise stabiliser is not contained in the Lie group of determinant-one matrices. \square

Now we are ready to prove Theorem 1.2, which is the second main theorem of the paper.

Proof of Theorem 1.2. The G -equivariance of T implies that $T(g \cdot \mathbf{x}_{(n)}) = g \cdot T(\mathbf{x}_{(n)})$ for every $g \in G$ and $\mathbf{x}_{(n)} \in (\mathbb{R}^m)^n$. In particular, the G^0 -equivariance implies that for every $\mathbf{x}_{(n)} \in (\mathbb{R}^m)^n$ the G^0 -stabilizer of $\mathbf{x}_{(n)}$ is contained in the G^0 -stabilizer of $T(\mathbf{x}_{(n)})$:

$$(21) \quad G_{\mathbf{x}_{(n)}}^0 \leq G_{T(\mathbf{x}_{(n)})}^0.$$

Note that since $T(\mathbf{x}_{(n)}) \in \mathcal{S}_{\mathcal{G}}^+$, its stabilizer is a generalized orthogonal group and hence it is a compact group. On the other hand, by (the proof of) Theorem 6.6, for a generic n -tuple $\mathbf{x}_{(n)}$ is not compact unless it is trivial, which happens only when $n \geq \max_{i \in [m]} |\downarrow i|$. Hence, if $n < \max_{i \in [m]} |\downarrow i|$ this necessarily contradicts (21) and then there are no G^0 -equivariant maps and in particular there are also no G -equivariant maps.

Now we show that $n \geq \max_{j \in [m]} |\downarrow j|$ is also a sufficient condition for the existence of an equivariant map T . It is sufficient to show this for $n = \max_{j \in [m]} |\downarrow j|$. In this case the G^0 -stabilizer of a generic point $\mathbf{x}_{(n)}$ is trivial and hence a generic G^0 -orbit in $(\mathbb{R}^m)^n$ has dimension $\dim G = \dim G^0 = |\{(i, j) : i \preccurlyeq j\}|$. The dimension of the ambient space is equal to mn . Let

$$d_i := n - |\downarrow i| \geq 0$$

for every $i \in [m]$. Then $\sum_i d_i$ is equal to the codimension of a generic G^0 -orbit in $(\mathbb{R}^m)^n$.

For each $i \in [m]$ fix a generic linear subspace L_i of \mathbb{R}^n of dimension d_i . Do it in such a way that L_i is the same for all i in a single orbit under the group of graph automorphisms. Let $L = \prod_{i \in [m]} L_i \subseteq (\mathbb{R}^n)^m$ then L can be identified with a subspace of $(\mathbb{R}^m)^n$ by transposition. For generic $\mathbf{x}_{(n)}$ the condition that the n -tuple $((g\mathbf{x}_1)_i, \dots, (g\mathbf{x}_n)_i)$ lies in L_i imposes $|\downarrow i|$ linear constraints on the i -th row of g . Hence there is a unique solution for that i -th row. It follows that L intersects a generic G^0 -orbit in a unique point.

Since L is stable under the group automorphism group $\text{Aut}(\mathcal{G})$ then we can consider the quotient $L/\text{Aut}(\mathcal{G})$. The induced projection from $(\mathbb{R}^m)^n$ onto $L/\text{Aut}(\mathcal{G})$ is denoted by π . By construction π is G -invariant. Take now any map $T_n : L/\text{Aut}(\mathcal{G}) \rightarrow \mathcal{S}_{\mathcal{G}}^+$. We extend the map for every $\mathbf{x}_{(n)} \in (\mathbb{R}^m)^n$ by taking $T_n(\mathbf{x}_{(n)}) := g \cdot T_n(\pi(\mathbf{x}_{(n)}))$ where $\mathbf{x}_{(n)} = g\pi(\mathbf{x}_{(n)})$. This map is defined almost everywhere and is G -equivariant by construction. Finally, the equation $\max_{i \in [m]} |\downarrow i| = \max_{A \in \mathbf{J}(\mathcal{C})} |A|$ follows directly from Lemma 4.3. \square

By Lemma 4.3 $Q = \max_{i \in [m]} |\downarrow i|$ is equal to the cardinality of the biggest element of $\mathbf{J}(\mathcal{C})$. For this reason we call Q the *irreducible maximal clique size*. If q is the maximal clique size then of course $q \geq Q$.

Proof of Theorem 2.6. Note that if a subgroup $H \subseteq G^0$ is a pointwise stabilizer of $B \subseteq \mathbb{R}^m$ then it is also a pointwise stabilizer of the smallest linear subspace containing B . Hence the subsets B in the definition of $\Delta(P)$ in (6) can be taken to be in the family of all linear subspaces of \mathbb{R}^m .

By Theorem 6.6 the maximal number k such that the set of k generic vectors has a pointwise stabilizer not consisting entirely of determinant-one matrices (hence $G^0 \cap G_1 \neq \emptyset$) equals $Q - 1$. This implies that the maximal dimension of a linear subspace with a pointwise stabilizer not consisting entirely of determinant-one matrices is $Q - 1$ as well. This gives the result by the definition of $\Delta(P)$ in (6). The last statement of the theorem follows from Theorem 2.4 and the fact that if $\mathbf{x}_{(n)}$ is generic then $\Delta(P_{\mathbf{x}_{(n)}}) = (Q - 1)/n$. \square

7. G -ORBITS IN THE CONE $\mathcal{S}_{\mathcal{G}}^+$

Given an undirected graph $\mathcal{G} = ([m], E)$ we have determined the group $G \subseteq \text{GL}_m(\mathbb{R})$ of all invertible linear maps $\mathbb{R}^m \rightarrow \mathbb{R}^m$ stabilizing the cone $\mathcal{S}_{\mathcal{G}}^+$, or, equivalently, stabilizing the vector space $\mathcal{S}_{\mathcal{G}}$ of symmetric matrices with zeroes on the off-diagonal entries corresponding to non-edges of \mathcal{G} . To complete our geometric description of the group G we need to understand the structure of G -orbits on $\mathcal{S}_{\mathcal{G}}^+$. The following result is important for statisticians as it tells us when the graphical Gaussian model forms a transformation family.

Theorem 7.1 (Theorem 2.2, [15]). *Let $\mathcal{G} = ([m], E)$ be an undirected graph. There is exactly one orbit of the G -action on $\mathcal{S}_{\mathcal{G}}^+$ if and only if for any two neighbors $i, j \in [m]$ either $i \preceq j$ or $j \preceq i$; or equivalently if \mathcal{G} does not contain $\bullet^i - \bullet^j - \bullet^k - \bullet^l$ as an induced subgraph.*

From the inferential point of view it is important to analyze also other cases, when the action of G on $\mathcal{S}_{\mathcal{G}}^+$ is not transitive and hence the model forms a composite transformation

family. In this section we answer the first natural question about the generic dimension of G -orbits on \mathcal{S}_G^+ and \mathcal{S}_G (this generic dimension, attained by symmetric matrices in some open dense subset, coincides on these two sets). Standard Lie group theory implies that this dimension equals $\dim G$ minus the dimension of the stabilizer of a generic matrix $K \in \mathcal{S}_G$. Thus we set out to determine the dimension of that stabilizer.

We do this in the following slightly larger generality. Let $\tilde{\mathcal{G}}$ denote the graph induced by \mathcal{G} on the set C of equivalence classes of the pre-order \preceq and let $c : C \rightarrow \mathbb{N}$, $i \mapsto c_i$ be the coloring mapping an equivalence class to its cardinality. We write $i \sim j$ to express that i and j are connected in $\tilde{\mathcal{G}}$. Let \leq be a partial order on C which is compatible with the edge set of $\tilde{\mathcal{G}}$ in the sense that $i \leq j \Rightarrow N(j) \cup \{j\} \subseteq N(i) \cup \{i\}$. Note that we do not require the implication \Leftarrow here, so \leq may have less comparable elements than the partial order on C coming from the pre-order \preceq . To every $i \in C$ we associate a vector space $U_i := \mathbb{R}^{c_i}$, and we set $U := \bigoplus_{i \in C} U_i$. We want to determine the (dimension of the) stabilizer of K in the group

$$H := \{g \in \mathrm{GL}(U) \mid gU_i \subseteq \bigoplus_{j \geq i} U_j\},$$

which, if \leq comes from \preceq , is nothing but the identity component G^0 of G . The final answer is as follows.

Theorem 7.2. *The dimension of the stabilizer H_K in H of a generic matrix in \mathcal{S}_G (or \mathcal{S}_G^+) equals $\sum_{i \in C} \binom{n_i}{2}$, where n_i is defined by*

$$n_i := \max \left\{ 0, c_i - \left(\sum_{j \sim i, i \not\leq j \not\leq i} c_j \right) \right\}.$$

In words: starting from $\tilde{\mathcal{G}}$, one deletes all edges between vertices that are comparable in the partial order \leq , and one subtracts from c_i the sum of the c_j for all neighboring j in the new graph. If the result is non-negative, then this is n_i ; otherwise, n_i is zero. The expression above suggests that the identity component of H_K is a product of special orthogonal groups of spaces of dimensions n_i , which is indeed what the proof will show. In the following proof we will identify K with the bilinear form on U defined by $K(\mathbf{x}, \mathbf{y}) := \mathbf{x}^T K \mathbf{y}$, and terms such as orthogonal complement and orthogonal group will refer to that form or restrictions of it.

Proof. We first prove the statement in the case where \leq is trivial, so that H consists of all linear maps $h \in \mathrm{GL}(U)$ that are block diagonal in the sense that they preserve the direct sum decomposition $U = \bigoplus U_i$. For each $i \in C$ let $U'_i \subseteq U_i$ denote the intersection with U_i of the orthogonal complement of $\bigoplus_{j \neq i} U_j$. Since the U_j with $j \not\sim i$ are perpendicular to U_i , the space U'_i also equals the space of vectors in U_i perpendicular to all spaces U_j with $j \sim i$. By genericity of K , these orthogonality conditions defining U'_i are as linearly independent as possible, and it follows that the dimension of U'_i is n_i . Given any linear map h in $\mathrm{O}(U'_i)$, i.e., preserving the restriction of K to U'_i , we extend it by the identity on the orthogonal complement of U'_i . The resulting map, still denoted h , stabilizes K by

construction, and we claim that it also maps each U_j into itself. For $j \neq i$ this is clear since U_j is perpendicular to U'_i and hence h is the identity on U_j ; for $j = i$ this is clear since h preserves U'_i and acts as the identity on the orthogonal complement of U'_i in U_i . Thus h lies in H_K . Doing this for all i yields a product of orthogonal groups, one for each U'_i , which is contained in H_K . The dimension of this product is the expression in the theorem. We claim that the identity component of H_K equals the identity component of this product (which is the product of special orthogonal groups).

To see this, we first discuss the case where $\tilde{\mathcal{G}}$ has two vertices connected by an edge. Label the vertices 1 and 2, with cardinalities c_1, c_2 , respectively. Without loss of generality we assume that $c_1 \geq c_2$. Then any $g \in H_K$ must stabilize U'_1 defined as above (note that U'_2 is zero), and hence also the orthogonal complement U''_1 of U'_1 in U_1 . Note that U''_1 and U_2 have the same dimension. The restriction of g to $U''_1 \oplus U_2$ is block diagonal with blocks $g_1 \in O(U''_1)$ and $g_2 \in O(U_2)$ and moreover satisfies $K(g_1 \mathbf{x}_1, g_2 \mathbf{x}_2) = K(\mathbf{x}_1, \mathbf{x}_2)$ for $\mathbf{x}_1 \in U''_1$ and $\mathbf{x}_2 \in U_2$. We claim that for generic K the group of such pairs (g_1, g_2) is zero-dimensional. It suffices to prove this for a specific K , and we take K , or rather its restriction to $U''_1 \oplus U_2$, to have the block structure

$$K = \begin{bmatrix} I & I \\ I & D \end{bmatrix}$$

where I is an identity matrix of size $\dim U''_1 = \dim U_2 = c_2$ and D is a diagonal matrix all of whose entries are distinct. Now the Lie algebra of the group of pairs (g_1, g_2) as above consists of all block diagonal matrices $A = \text{diag}(A_1, A_2)$ of $c_2 \times c_2$ -matrices satisfying $A^T K + K A = 0$. Substituting the blocks yields

$$\begin{bmatrix} A_1^T + A_1 & A_1^T + A_2 \\ A_2^T + A_1 & A_2^T D + D A_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

This means that A_1 is skew-symmetric (in the ordinary sense, i.e., its (i, j) -entry is minus its (j, i) -entry), and hence that $A_2 = -A_1^T = A_1$. But A_2 also satisfies $(A_2)_{ji} D_{jj} = -D_{ii} (A_2)_{ij}$. Using that the D_{ii} are distinct one finds that $A_2 = A_1 = 0$. Hence said Lie algebra is zero, and the identity component of the group of pairs (g_1, g_2) is trivial, hence the identity component of H_K reduces to $\text{SO}(U'_1)$, as claimed.

Next we discuss case where $\tilde{\mathcal{G}}$ is general but \leq is still trivial. As mentioned earlier, this means that H consists of block diagonal linear maps; denote by h_i the restriction of $h \in H$ to U_i . We want to prove that for any h in the identity component of H_K , each h_i lies in $\text{SO}(U'_i)$ in the sense that it is the identity on the orthogonal complement of U'_i inside U_i . For each $j \sim i$ the pair (h_i, h_j) stabilizes the restriction K' of K to $U_i \oplus U_j$, and lies in $H'_{K'}$, where $H' \subseteq H$ is the subgroup $\text{GL}(U_i) \times \text{GL}(U_j)$. By the discussion of the two-vertex case above, we find that h_i is the identity on the orthogonal complement in U_i of $U_i \cap U_j^\perp$. Note that, by genericity of K , this orthogonal complement is all of U_i if $c_i \leq c_j$, and of dimension $c_i - (c_i - c_j) = c_j$ otherwise. Varying j over the neighbours of i , we find that h_i is the identity on the sum over j of all these orthogonal complements, which is the same thing as the orthogonal complement in U_i of $\bigcap_{j \sim i} (U_i \cap U_j^\perp) = U'_i$. This proves the theorem in the case where \leq is trivial.

It remains to prove the theorem in the case where \leq is non-trivial. Then any $h \in H$ stabilizes the spaces $U_{\geq i} := \bigoplus_{j \geq i} U_j$ and $U_{> i} := \bigoplus_{j > i} U_j$. If h also stabilizes K , then it stabilizes the orthogonal complement $V_i := U_{\geq i}^\perp \cap U_{\geq i}$ of $U_{> i}$ inside $U_{\geq i}$. Note that V_i has dimension c_i , and we argue that we can replace the U_i by the V_i . Indeed, for any pair $i, j \in C$ consider the restriction of the bilinear form K to $V_i \times V_j$. If $i \not\sim j$, then we claim that this restriction is zero. Indeed, each element of V_i is a linear combination of elements of the spaces $U_{i'}$ with $i' \geq i$, and similarly for the elements of V_j . Now if some $i' \geq i$ were connected to some $j' \geq j$, then by the compatibility of the partial order with the edge set we find that also $i \sim j'$ (since $i \leq i'$) and hence $i \sim j$ (since $j \leq j'$), a contradiction. Hence no $i' \geq i$ is connected to any $j' \geq j$, and therefore the restriction of K to $U_{\geq i} \times U_{\geq j}$ is zero. In particular, K restricts to zero on $V_i \times V_j$. Next, if $i \sim j$ but i and j are comparable in the partial order, then by construction V_i is perpendicular to V_j . Hence only if $i \sim j$ and i, j are *not* comparable can K restrict to something non-zero (and generic) on $V_i \times V_j$. Thus upon replacing the U_i by the V_i and deleting from $\tilde{\mathcal{G}}$ the edges among vertices comparable in the partial order we are reduced to the previous case, where \leq is trivial. This concludes the proof of the dimension formula. \square

Note in particular that Theorem 7.2 confirms that, in the situation of Theorem 7.1, the dimension of the stabilizer of a generic matrix in $\mathcal{S}_{\mathcal{G}}^+$ is $\sum_{i \in C} \binom{c_i}{2}$. Consequently, the dimension of a generic orbit equals $\dim G$ minus this expression, which under the assumption of Theorem 7.1 is readily seen to equal the dimension of the model, in accordance with the fact that $\mathcal{S}_{\mathcal{G}}^+$ is a single orbit. Another example is given by the bull graph.

Example 7.3. Consider the bull graph in Figure 4 but with each vertex now representing an equivalence class in C with cardinality c_i for $i = 1, \dots, c_5$. In this case the only pair of connected but not comparable vertices is $(1, 2)$. With the convention that $\binom{m}{2} = 0$ if $m \leq 0$, Theorem 7.2 gives that the dimension of the stabilizer of a generic matrix in $\mathcal{S}_{\mathcal{G}}^+$ is

$$\binom{c_1 - c_2}{2} + \binom{c_2 - c_1}{2} + \binom{c_3}{2} + \binom{c_4}{2} + \binom{c_5}{2}.$$

Example 7.4. Let now $\tilde{\mathcal{G}}$ be a tree, where each vertex v represents an equivalence class with cardinality c_v . In this case the dimension of the stabilizer of a generic matrix in $\mathcal{S}_{\mathcal{G}}^+$ is

$$\sum_{(u,v) \in \text{inner}} \left(\binom{c_u - c_v}{2} + \binom{c_v - c_u}{2} \right) + \sum_{i \in \text{leaves}} \binom{c_i}{2},$$

where the first sum is over all the inner vertices of $\tilde{\mathcal{G}}$ and the second sum is over all the leaves (vertices of valency 1) of $\tilde{\mathcal{G}}$. In particular if for some c we have that $c_i = c$ for all $i \in C$ then this formula degenerates to $l \binom{c}{2}$, where l is the number of leaves.

8. DISCUSSION

Given an undirected graph $\mathcal{G} = ([m], E)$ we have determined the group $G \subseteq \text{GL}_m(\mathbb{R})$ of all invertible linear maps $\mathbb{R}^m \rightarrow \mathbb{R}^m$ stabilizing the vector space $\mathcal{S}_{\mathcal{G}}$ of symmetric matrices

with zeroes on the off-diagonal entries corresponding to non-edges of \mathcal{G} . This helped us to analyze bounds on the finite sample breakdown points for G -equivariant estimators for Gaussian graphical models and generic data sets. Note that in practice often $\mathbf{x}_{(n)}$ is not generic. Consider for example the problem of modeling the height of people with measurements taken in centimeters. In a typical sample for this case we can expect many repeated data points. In this case a linear subspace of dimension $Q - 1$ may contain more than $Q - 1$ data points which will lead to higher values of $\Delta(P_{\mathbf{x}_n})$. It would be helpful to have efficient ways of computations for non-generic data sets. Note also that the proof of Theorem 1.2 gives a general construction of a G -equivariant estimator. It would be interesting to see how this construction helps to construct robust estimators with some explicit high lower bounds on the finite sample breakdown points.

APPENDIX A. MATRIX GROUPS AND LIE ALGEBRAS

In this appendix we collect some basic facts about Lie groups that are used in the paper. For details see for example [3]. We only need the notion of a *matrix Lie group*, which is a closed subgroup G of the group $\mathrm{GL}_m(\mathbb{R})$ of invertible real $m \times m$ -matrices. It follows that G is, in fact, a smooth submanifold of $\mathrm{GL}_m(\mathbb{R})$. Two examples featuring in the paper are the group $\mathrm{SL}_m(\mathbb{R})$ consisting of determinant-one matrices and the generalized orthogonal groups $\mathrm{O}_K(\mathbb{R})$, one for every positive definite symmetric matrix K , consisting of matrices g such that $g^T K g = K$ (we use notation $\mathrm{O}(\mathbb{R})$ if K is generic - see Section 7). Among these, $\mathrm{SL}_m(\mathbb{R})$ is connected, while $\mathrm{O}_K(\mathbb{R})$ has two connected components, distinguished by the value ± 1 of the determinant. The groups $\mathrm{O}_K(\mathbb{R})$ are compact, because they are closed and bounded in the Euclidean norm on the space of $m \times m$ -matrices (for this one needs the fact that K is positive definite). Two further classes of examples are T^m , the group of diagonal invertible $m \times m$ -matrices, and the symmetry groups of Gaussian graphical models discussed in the paper.

To understand a matrix Lie group G it helps to understand its tangent space at the identity matrix. This vector space of $m \times m$ -matrices is usually denoted \mathfrak{g} , and has the structure of a *matrix Lie algebra*, i.e., it is closed under taking matrix commutators. A basic fact used in the paper is that \mathfrak{g} uniquely determines the connected component G^0 of G that contains the identity matrix. Using the fact that the derivative of the determinant function at the identity matrix is the trace, one finds that the Lie algebra of $\mathrm{SL}_m(\mathbb{R})$ consists of all trace-zero matrices. Similarly, by taking derivatives of the equations defining $\mathrm{O}_K(\mathbb{R})$ one finds that its Lie algebra consists of all matrices A satisfying $A^T K + K A = 0$.

The following well-known lemma connects certain matrix Lie groups to the combinatorics of preorders, and plays a central role in the paper. For completeness, we give a proof.

Lemma A.1. *Let $H \subseteq \mathrm{GL}_m(\mathbb{R})$ be a matrix Lie group containing the group T^m . Then the Lie algebra of H has a basis consisting of matrices E_{ij} with (i, j) running through some subset I of $[m] \times [m]$. Moreover, the set I defines a preorder on $[m]$ in the sense that (i, i) lies in I for all $i \in [m]$ and $(i, j), (j, k) \in I \Rightarrow (i, k) \in I$. Conversely, the E_{ij} with (i, j) running through any set $I \subseteq [m] \times [m]$ defining a pre-order span the Lie algebra of a closed subgroup of $\mathrm{GL}_m(\mathbb{R})$.*

Proof. Let $\mathfrak{t} \simeq \mathbb{R}^m$ denote the Lie algebra of T^m , which is spanned by the E_{ii} for $i \in [m]$. Then \mathfrak{t} is a Lie subalgebra of \mathfrak{h} . Let $A = \sum_{i \neq j} a_{ij} E_{ij}$ be any element of \mathfrak{h} with zero entries on the diagonal. We show that $E_{kl} \in \mathfrak{h}$ for every (k, l) such that $a_{kl} \neq 0$. We have

$$[E_{ll}, [E_{kk}, A]] = -a_{lk} E_{lk} - a_{kl} E_{kl} \in \mathfrak{h}.$$

If only one of a_{kl} and a_{lk} is non-zero, then the result follows immediately. If both are non-zero, then note that

$$[E_{kk}, [E_{ll}, [E_{kk}, A]]] = a_{lk} E_{lk} - a_{kl} E_{kl}.$$

Since \mathfrak{h} is a linear subspace, we conclude that both E_{kl} and E_{lk} lie in \mathfrak{h} .

Thus \mathfrak{h} is spanned by matrices E_{ij} with (i, j) running through some subset I of $[m] \times [m]$. The inclusion $\mathfrak{t} \subseteq \mathfrak{h}$ implies that $(i, i) \in I$ for all i . Now if $E_{ij}, E_{jk} \in \mathfrak{h}$ with $k \neq i$, then $[E_{ij}, E_{jk}] = E_{ik}$ so that $(i, k) \in I$, as well. This yields the preorder property of I .

Finally, if $I \subseteq [m] \times [m]$ determines a pre-order, then the set of all invertible matrices g with $g_{ij} = 0$ unless $(i, j) \in I$ form a matrix Lie group (by a computation similar to the one just performed), whose Lie algebra is spanned by the E_{ij} with $(i, j) \in I$. \square

In the paper we also implicitly make use of the fact that the matrix exponential map maps the Lie algebra of a matrix Lie group into the group, but only in the basic case where $\exp(tE_{ij}) = 1 + tE_{ij}$.

REFERENCES

- [1] S. A. ANDERSSON, D. MADIGAN, M. D. PERLMAN, AND C. M. TRIGGS, *On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs*, J. Statist. Plann. Inference, 48 (1995), pp. 25–46.
- [2] S. A. ANDERSSON AND M. D. PERLMAN, *Lattice models for conditional independence in a multivariate normal distribution*, Ann. Statist., 21 (1993), pp. 1318–1358.
- [3] A. BAKER, *Matrix groups*, Springer Undergraduate Mathematics Series, Springer-Verlag London Ltd., London, 2002. An introduction to Lie group theory.
- [4] O. BARNDORFF-NIELSEN, P. BLÆSILD, J. L. JENSEN, AND B. JØRGENSEN, *Exponential transformation models*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 379 (1982), pp. 41–65.
- [5] C. BECKER, *Iterative proportional scaling based on a robust start estimator*, in Classification, the ubiquitous challenge: proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation eV, University of Dortmund, March 9–11, 2004, vol. 28, Springer-Verlag New York Inc, 2005, p. 248.
- [6] M. CHUDNOVSKY, *The structure of bull-free graphs i—three-edge-paths with centers and anticenters*, Journal of Combinatorial Theory, Series B, 102 (2012), pp. 233 – 251.
- [7] P. DAVIES AND U. GATHER, *The breakdown point—examples and counterexamples*, REVSTAT—Statistical Journal, 5 (2007), pp. 1–17.
- [8] P. L. DAVIES AND U. GATHER, *Breakdown and groups*, The Annals of Statistics, 33 (2005), pp. 977–988.
- [9] M. FINEGOLD AND M. DRTON, *Robust graphical modeling of gene networks using classical and alternative t-distributions*, The Annals of Applied Statistics, 5 (2011), pp. 1057–1080.
- [10] R. A. FISHER, *Two new properties of mathematical likelihood*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 144 (1934), pp. 285–307.

- [11] A. GOTTARD AND S. PACILLO, *On the impact of contaminations in graphical gaussian models*, Statistical Methods & Applications, 15 (2007), pp. 343–354.
- [12] A. GOTTARD AND S. PACILLO, *Robust concentration graph model selection*, Computational Statistics & Data Analysis, 54 (2010), pp. 3070–3079.
- [13] F. R. HAMPEL, *A general qualitative definition of robustness*, Ann. Math. Statist., 42 (1971), pp. 1887–1896.
- [14] S. KUHN AND C. BECKER, *Sensitivity of graphical modeling against contamination*, in Between Data Science and Applied Data Analysis, M. Schader, W. Gaul, and M. Vichi, eds., Studies in Classification, Data Analysis, and Knowledge Organization, Springer Berlin Heidelberg, 2003, pp. 279–287.
- [15] G. LETAC AND H. MASSAM, *Wishart distributions for decomposable graphs*, Ann. Statist., 35 (2007), pp. 1278–1323.
- [16] R. A. MARONNA, R. D. MARTIN, AND V. J. YOHAI, *Robust statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., Chichester, 2006. Theory and methods.
- [17] N. MEINSHAUSEN AND P. BÜHLMANN, *High-dimensional graphs and variable selection with the lasso*, The Annals of Statistics, 34 (2006), pp. 1436–1462.
- [18] M. MIYAMURA AND Y. KANO, *Robust gaussian graphical modeling*, Journal of multivariate analysis, 97 (2006), pp. 1525–1550.
- [19] P. D. RAVIKUMAR, G. RASKUTTI, M. J. WAINWRIGHT, AND B. YU, *Model selection in gaussian graphical models: High-dimensional consistency of l_1 -regularized mle.*, in NIPS, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds., Curran Associates, Inc., 2008, pp. 1329–1336.
- [20] M. J. SCHERVISH, *Theory of statistics*, Springer Series in Statistics, Springer-Verlag, New York, 1995.
- [21] R. P. STANLEY, *Enumerative combinatorics. Volume I*, no. 49 in Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2002.
- [22] C. UHLER, *Geometry of maximum likelihood estimation in Gaussian graphical models*, The Annals of Statistics, 40 (2012), pp. 238–261.
- [23] D. VOGEL AND R. FRIED, *Elliptical graphical modelling*, Biometrika, 98 (2011), pp. 935–951.

TU EINDHOVEN, DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, PO Box 513, 5600 MB EINDHOVEN, AND CWI AMSTERDAM, THE NETHERLANDS
E-mail address: `j.draisma@tue.nl`

TU DORTMUND UNIVERSITY, FACULTY OF STATISTICS, 44221 DORTMUND, GERMANY
E-mail address: `kuhn@statistik.tu-dortmund.de`

TU EINDHOVEN, DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, PO Box 513, 5600 MB EINDHOVEN, THE NETHERLANDS
E-mail address: `piotr.zwiernik@gmail.com`